

# The Volatility of Long-Term Bond Returns

Citation for published version (APA):

Osterrieder, D., & Schotman, P. (2017). The Volatility of Long-Term Bond Returns: Persistent Interest Shocks and Time-Varying Risk Premiums . *Review of Economics and Statistics*, 99(5), 884-895.  
[https://doi.org/10.1162/REST\\_a\\_00624](https://doi.org/10.1162/REST_a_00624)

**Document status and date:**

Published: 01/12/2017

**DOI:**

[10.1162/REST\\_a\\_00624](https://doi.org/10.1162/REST_a_00624)

**Document Version:**

Publisher's PDF, also known as Version of record

**Document license:**

Taverne

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

**Take down policy**

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

# THE VOLATILITY OF LONG-TERM BOND RETURNS: PERSISTENT INTEREST SHOCKS AND TIME-VARYING RISK PREMIUMS

Daniela Osterrieder and Peter C. Schotman\*

**Abstract**—We develop an almost affine term-structure model with a closed-form solution for factor loadings in which the spot rate and the risk price are fractionally integrated processes with different integration orders. This model is used to explain two stylized facts. First, predictability of long-term excess bond returns requires sufficient volatility and persistence in the risk price. Second, the large volatility of long-term bond returns requires persistence in the spot rate. Decomposing long-term bond returns, we find that the expectations component from the level factor is more volatile than returns themselves and that the risk premium correlates negatively with level-factor innovations.

## I. Introduction

**A**FFINE term structure models assume a stationary vector autoregressive (VAR) process for the factors that drive interest rates. Time series estimates of the VAR typically imply that long-term expectations of future spot rates have very little variation. With constant risk premiums, long-term yields would then have very low volatility. In the data, long-term yields are, however, almost as volatile as short-term yields. To explain this volatility in the data, risk premiums need to be very volatile. This implication hinges on the estimated persistence in the VAR model.<sup>1</sup> To explain the volatility puzzle without highly volatile risk premiums, mean reversion in interest rates must be very slow, and hence the VAR models need a near unit root.

The sensitivity with respect to near unit root parameters is illustrated in Cochrane and Piazzesi (2008) and Jardet, Monfort, and Pegoraro (2013). Both studies compare the long-run forecasts of the short-term interest rate from estimated VAR models with and without imposing cointegration among yields of different maturities. These long-run forecasts are very different. For the stationary VAR, they are close to the constant unconditional mean of the spot rate, whereas the cointegrated model produces forecasts that are very close to the current level of the spot rate.

Received for publication August 3, 2012. Revision accepted for publication June 14, 2016. Editor: Mark W. Watson.

\* Osterrieder: Rutgers Business School and CREATES; Schotman: Maastricht University and NETSPAR.

We thank participants of the Quantitative Methods in Finance Conference in Sydney, the SOFIE Meeting in Singapore, and the 'Forecasting Structure and Time Varying Patterns in Economics and Finance' in Rotterdam for helpful comments. We also thank Antoon Pelsser, Jean-Pierre Urbain, Rolf Tschernig, Andrew Harvey, John Geweke, and Stefan Straetmans for their suggestions. This research was financially supported by a grant of the Dutch Organization for Scientific Research (NWO), and by CREATES—Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

<sup>1</sup> See Bauer (2011) for an analysis of the relation between the excess volatility puzzle and persistent interest rate shocks. The puzzle has a long history. Shiller (1979) already conjectured a link between volatility and unit roots: "We have no real information in small samples about possible trends or long cycles in interest rates. Indeed, . . . short-term interest rates may be nonstationary" (p. 1213).

How to deal with this sensitivity remains problematic. Estimates of the largest autoregressive root are biased downward due to the well-known Kendall bias, which is exacerbated in multivariate systems (see Abadir, Hadri, & Tzavalis, 1999). Recently various bias-adjustment procedures have been proposed.<sup>2</sup> Instead of further refining the estimators of Gaussian affine models, we extend the class of models by including fractionally integrated processes and breaking the affine relation between the risk premium and the level of state variables.

Fractional integration, denoted  $I(d)$ , is a parsimonious and flexible means to model the long-memory properties of interest rate dynamics, as it allows a smooth transition between stationary  $I(0)$  processes and nonstationary unit-root  $I(1)$  processes. The fractional model can generate forecasts that are in between the stationary and cointegrated models. We see three motivations for applying fractional integration to model the term structure of interest rates. First, many studies have estimated the fractional integration parameter of interest rate time series and report that the order  $d$  is between 0.8 and 1 but significantly different from 0.<sup>3</sup> For our sample of 58 years of monthly observations, we estimate the order of integration as  $d = 0.89$ . That means that the level of interest rates is nonstationary but less persistent than a random walk or  $I(1)$  process.

Second, fractional integration models are linear and therefore analytically tractable. We obtain an analytical solution for the term structure for any Gaussian linear process for the spot rate combined with any Gaussian linear process for the price of risk, jointly driven by  $K$  shocks. Since the price of risk is not necessarily related to the level of the state variables, risk premiums can be stationary even though state variables are not. In the model, excess returns have a factor structure with factor loadings given in closed form. For this reason, we prefer the fractional model to other model classes that can also generate long-memory-like behavior, such as regime-switching models. As Diebold and Inoue (2001) have shown, a fractionally integrated model provides a good approximation for long-run predictions for time series that are subject to occasional breaks in the mean. Connolly, Güner, and Hightower (2007) further demonstrate that

<sup>2</sup> Bauer, Rudebusch, and Wu (2012) develop a bootstrap adjustment of the VAR parameters. Jardet et al. (2013) take a weighted average of the stationary and cointegrated forecasts. Joslin, Priebsch, and Singleton (2014) restrict the largest eigenvalues under the physical  $\mathbb{P}$  and risk-neutral  $\mathbb{Q}$  measure to be equal. Dewachter and Lyrio (2006) impose the unit root under  $\mathbb{P}$ , whereas Christensen, Diebold, and Rudebusch (2011) impose the unit root under  $\mathbb{Q}$ . Cochrane and Piazzesi (2008) estimate persistence under  $\mathbb{Q}$  and infer the persistence under  $\mathbb{P}$  by constraining the specification for risk prices.

<sup>3</sup> See, e.g., Shea (1991), Sun and Phillips (2004), Iacone (2009), and Gil-Alana and Moreno (2012).

a long-memory model for the short rate may describe the series more accurately than a structural change model.

Third, fractional models have been shown to fit certain characteristics of the cross section. Our approach has been motivated by Backus and Zin (1993), who assumed an autoregressive fractionally integrated moving average, ARFIMA( $p, d, q$ ), process for the spot rate, together with a constant risk premium. With this model, they succeed in matching the observed mean and volatility of yields. We relax two of their assumptions. First, we do not require that interest rates be stationary with  $d < 1/2$ . The assumption was necessary for their tests of unconditional moments of yield levels, but it is not required for our tests on the volatility of excess returns. Second, we allow for time-varying risk premiums because there is strong empirical evidence that excess returns have a predictable component and because these may be an important source of volatility in bond returns.

For the price of risk, we also consider a fractional model. A large literature relates the predictable part of excess returns to yield spreads or forward premiums. Campbell and Shiller (1991) and many others generally report significant predictive power.<sup>4</sup> Following Campbell and Shiller (1987), cointegration models of the term structure assume that spreads are  $I(0)$ . More recent time series tests reach a different conclusion. Both Chen and Hurvich (2003) and Nielsen (2010) find that spreads have a fractional order that is significantly larger than 0, but with point estimates that are less than a half. We find the same in our data, not only in time series tests but also implied by stylized facts at the long end of the term structure.

A related approach is the shifting end points model of Kozicki and Tinsley (2001). They specify long-horizon expectations of the short-term interest rate as a (nonlinear) function of inflation and inflation expectations. The shifting end point serves as the level factor for the term structure and is successful in tracking long-maturity yields. The long-memory properties of the factor are due to the nonstationarity of inflation and in some specifications to regime shifts in the monetary policy target inflation rate. We interpret their evidence as an indication that fractional models may provide the right amount of persistence.

Our focus is on long-maturity bonds. At long maturities, only the low-frequency components in the level factor matter, and we can work with a single-factor model as in Backus and Zin (1993) and Kozicki and Tinsley (2001). When we compare term structure estimates for different orders of integration, we find a strong interaction between the persistence of the level factor and the risk premium in long-maturity bond returns. With low persistence in the level factor, as in a stationary AR model, the correlation between risk prices and the spot rate is positive. With stronger persistence, such as our estimated  $d = 0.89$  or an  $I(1)$  unit root, the correlation

changes sign. In this case, the expectations-driven part of the volatility of excess returns is larger than the total volatility. This implies that the covariance between risk premiums and changes in expectations must be negative to match the volatility in the data. A similar sign change for the covariance between the spot rate and the risk premium is observed in Bauer, Rudebusch, and Wu (2014) in their comment on Wright (2011). In Wright (2011) the spot rate dynamics are stationary resulting in very stable long-term expectations and high implied volatility in the risk premium. Bauer et al. (2014) advocate a bias adjustment of the dynamics toward the unit root and as a result find that expectations become more volatile and that the covariance between the risk premium and macroeconomic variables changes sign.

## II. Linear Time Series Processes and the Term Structure

Our model is a generalization of the discrete-time, Gaussian, essentially affine model for the term structure of interest rates developed by Duffee (2002). Two differences are important: (a) the price of risk is not necessarily an affine function of the state variables and (b) the dynamics for the short-term interest rate can be more general than a VAR.

We assume that the one-period spot rate,  $r_t$ , can be represented by the linear moving average specification,

$$\Delta r_t = \sum_{j=0}^{\infty} c_j' \epsilon_{t-j}, \quad (1)$$

where  $c_j$  and  $\epsilon_t$  are vectors of length  $K$  and  $\epsilon_t \stackrel{iid}{\sim} N(0, \Sigma)$ . Depending on the assumptions on the impulse responses  $c_j$ , the general formulation in equation (1) encompasses many time series models, including models with fractional integration in which the spot rate is integrated of order  $I(d_r)$ . The model is formulated in first differences to allow for non-stationary processes. For a stationary process, the model is overdifferenced, and information on the unconditional mean of  $r_t$  is lost. In this paper, we will focus on excess returns and yield spreads. Both do not depend on the unconditional mean of the spot rate.

The logarithmic stochastic discount factor  $m_{t+1} = \ln M_{t+1}$  is specified as

$$m_{t+1} = -r_t - \frac{1}{2} \lambda_t' \Sigma \lambda_t + \lambda_t' \epsilon_{t+1}, \quad (2)$$

where  $\lambda_t$  is a  $K$ -vector of risk prices following the linear process

$$\Sigma \lambda_t = \Sigma \mu_\lambda + \sum_{j=0}^{\infty} F_j' \epsilon_{t-j}, \quad (3)$$

with  $\mu_\lambda$  a  $K$ -vector and  $F_j$  matrices of coefficients of order  $(K \times K)$ . The order of integration  $I(d_\lambda)$  of  $\lambda_t$  can be different from that of the spot rate. The specification is in levels, as we will always assume  $d_\lambda < 1/2$ .

<sup>4</sup> See, e.g., Fama (1984), Fama and Bliss (1987), Duffee (2002), and Cochrane and Piazzesi (2005).

Specifications (1) to (3) reduce to an essentially Gaussian affine term structure model if both the spot rate and the risk prices are affine in  $K$  state variables,  $X_t$ ,

$$r_t = \delta_0 + \delta'_1 X_t, \\ \Sigma \lambda_t = \Lambda_0 + \Lambda_1 X_t,$$

with state variables that follow the first-order VAR  $X_t = \mathcal{A}X_{t-1} + \epsilon_t$  (possibly with unit roots in  $\mathcal{A}$ ). In the affine model, the coefficients  $c_j$  and  $F_j$  are both determined by the VAR coefficient matrix  $\mathcal{A}$ . In the current model,  $c_j$  and  $F_j$  can be unrelated and do not need to generate the same time series properties. We will consider a model where the spot rate is a nonstationary process, whereas the price of risk remains stationary. In particular, we will consider a model in which the spot rate level is fractionally integrated of order  $d_r > 1/2$  and the price of risk integrated of order  $d_\lambda < 1/2$ .

Prices of discount bonds of maturity  $n$  are denoted  $P_t^{(n)}$ , log bond prices are  $p_t^{(n)} = \ln P_t^{(n)}$ , and continuously compounded yields are  $y_t^{(n)} = -(1/n)p_t^{(n)}$ . Bond prices satisfy the pricing equation

$$P_t^{(n+1)} = E_t \left[ M_{t+1} P_{t+1}^{(n)} \right], \quad (4)$$

with initial condition  $P_t^{(1)} = e^{-r_t}$ . Given spot rates and risk prices that are governed by equations (1) to (3), the dynamics for prices and yields of all maturities can be derived recursively from equation (4). Since we allow for nonstationary spot rate dynamics, it is more convenient to express yields relative to the spot rate level using the spread  $s_t^{(n)} = y_t^{(n)} - r_t$ . The implied dynamic process for the spread is summarized in theorem 1 (the proof is in appendix A):

**Theorem 1.** *The yield spread for a discount bond of maturity  $n$  follows the process*

$$s_t^{(n)} = \frac{1}{n} \left( a^{(n)} + \sum_{j=0}^{\infty} d_j^{(n)'} \epsilon_{t-j} \right) \quad (5)$$

with recursively defined coefficients

$$d_j^{(n+1)} = d_{j+1}^{(n)} + nc_{j+1} + F_j'(d_0^{(n)} + nc_0), \quad (6)$$

$$a^{(n+1)} = a^{(n)} - \frac{1}{2}(d_0^{(n)} + nc_0)' \Sigma (d_0^{(n)} + nc_0) \\ + (d_0^{(n)} + nc_0)' \Sigma \mu_\lambda, \quad (7)$$

and initial conditions  $a^{(1)} = 0$  and  $d_j^{(1)} = 0$ .

Given our assumptions on  $d_r$  and  $d_\lambda$ , theorem 1 implies that  $s_t^{(n)} \sim I(d_\lambda)$ . Our empirical analysis will be based on the excess returns defined as

$$rx_{t+1}^{(n+1)} \equiv p_{t+1}^{(n)} - p_t^{(n+1)} - r_t \\ = s_t^{(n+1)} - n \left( s_{t+1}^{(n)} - s_t^{(n+1)} + \Delta r_{t+1} \right), \quad (8)$$

where the second equality follows directly from the definition of the spread. For an explicit solution we use equation (5) with the coefficients in equation (6). The result is in theorem 2 (see appendix B).

**Theorem 2.** *Let the spot rate be generated by the linear process (1) and risk prices be generated by the linear process (3). Then excess returns on discount bonds have the factor structure*

$$rx_{t+1}^{(n+1)} = b_n' (-\epsilon_{t+1} + \Sigma \lambda_t) - \frac{1}{2} b_n' \Sigma b_n, \quad (9)$$

with factor loadings obeying  $b_1 = c_0$ , and

$$b_n = C_{n-1} + \sum_{i=1}^{n-1} F'_{n-1-i} b_i, \quad n > 1, \quad (10)$$

with  $C_{n-1} = \sum_{i=0}^{n-1} (n-i)c_i$  the cumulative impulse responses for the level of the spot rate.

Excess returns have three components. First, shocks  $\epsilon_{t+1}$  enter with factor loadings  $b_n$ . Recursion (10) for these loadings includes both the effects of the short rate process through the  $c_j$  terms, as well as the risk price dynamics through the  $F_j$  terms. The second element is the predicted excess return, which is linear in  $\Sigma \lambda_t$  with the same factor loading  $b_n$ . The last term is the Jensen inequality adjustment  $(1/2)b_n' \Sigma b_n$ . The general structure, equation (9), is the same as for the Gaussian essentially affine model. What is different are the coefficients  $b_n$ .<sup>5</sup> The second term in equation (10) shows how the time series process of risk prices affects the factor loadings. For long-term maturities, the relative magnitudes of the expectations effect  $C_{n-1}$  and the risk premium effect crucially depend on the rate of decline of the  $C_n$  sequence for the spot rate and the sign of the  $F_n$  impulses of the risk premium.

In this model, it is much more convenient to work with excess returns than yield levels or spreads. With excess returns, theorem 2 implies that we still maintain a low-dimensional factor structure, whereas the spreads in equation (5) do not allow any factor structure for general linear time series processes.

Our interest is in the effect of the spot-rate persistence on the volatility and predictability of excess returns on long-term bonds. For this analysis, we make two related assumptions. First, we will concentrate our analysis on long-maturity bond returns, since that is where we should expect to see the largest effects of alternative estimates of persistence. Second, we will focus on the most persistent term structure factor, known as the level factor. In the remainder, we therefore specialize our model to the single factor case, with the level factor as the only one. Adding factors with

<sup>5</sup> In the Gaussian essentially affine model, the two terms in the factor loadings collapse to a single term with risk-neutral dynamics:  $b_n = \bar{C}_{n-1}$  with  $\bar{c}_0 = \delta'_1$ ,  $\bar{c}_j = (\bar{\mathcal{A}}^{j-1}(\bar{\mathcal{A}} - I)\delta_1)'$  and  $\bar{\mathcal{A}} = \mathcal{A} + \Lambda_1$ .

TABLE 1.—SUMMARY STATISTICS

	Average	S.D.	1st AC	EW	LW
Three-months T Bill (level) $r_t$	0.404	0.248	0.984	0.926	0.900
Three-months T Bill (first difference) $\Delta r_t$	0.000	0.040	0.127		
Excess return five-year bond $rx_t^{(k)}$	0.102	1.809	0.099		
Excess return ten-year bond $rx_t^{(m)}$	0.129	3.036	0.071		
Five-year yield spread $s_t^{(k)}$	0.093	0.083	0.923	0.422	0.432
Ten-year yield spread $s_t^{(m)}$	0.123	0.110	0.947	0.541	0.563

Averages and standard deviations are expressed in percent per month. 1st AC is the first-order autocorrelation. EW and LW are the semiparametric exact Whittle (EW) and local Whittle (LW) estimates for the fractional differencing parameter  $d$  (see, e.g., Shimotsu & Phillips, 2006). For the LW estimator, the time series  $r_t$  is prefiltered by differencing it once and adding 1 back to the parameter estimate later. For the EW estimator, a correction for the unknown mean is applied by subtracting the first observation for the spot rate and the time series average for the spreads (Shimotsu, 2010). The size of the spectral window is  $T^{0.5} = 26$ . The asymptotic standard error in all cases equals  $\hat{s}(\hat{d}) = 0.0981$ .

less persistence than the level factor allows for flexible yield curve shapes at short and intermediate maturities, but such factors have negligible effects at the long end of the yield curve. According to Cochrane and Piazzesi (2005, 2008), the level factor is also responsible for most of the volatility in the risk premium.

### III. Data and Stylized Facts

We use monthly time series of U.S. long-term interest rates for 1954:01–2012:02. Zero-coupon yields ( $y_t^{(n)}$ ) with maturities of five and ten years are from the replication data<sup>6</sup> of Campbell and Viceira (2001) up to 1994:11, supplemented by the Gürkaynak, Sack, and Wright (2007) data<sup>7</sup> from 1994:12 onward. For the spot rate  $r_t$ , we use the three-month Treasury bill rate from the FRED database,<sup>8</sup> transformed to continuously compounded yields. To construct excess returns, we use the long maturity approximation  $rx_{t+1}^{(n+1)} = -n\Delta y_{t+1}^{(n)} + y_t^{(n)} - r_t$ .

The choice of maturities is motivated by two considerations. Both maturities should be long enough not to be affected by transitory factors, while the two maturities should be far enough apart to obtain an accurate estimate of the rate of increase of the factor loadings. Maturities longer than ten years would be even more informative, but for these, we do not have long time series data of liquid market prices. In the remainder, we will denote the two maturities by  $k$  (60 months) and  $m$  (120 months).

Summary statistics are in table 1. The key moment we wish to match is the relative volatility of five- and ten-year excess returns,  $\mathcal{M}_0$ . Using the sample standard deviations, we find  $\mathcal{M}_0 = 1.68$ . To interpret this number, suppose the spot rate were a random walk ( $c_j = 0$  for  $j > 0$ ) and the price of risk a constant. In that case, theorem 2 implies the factor loadings  $b_n = n$  and, hence, a volatility ratio

$\mathcal{M}_0 = b_m/b_k = 2$ . Conversely, if the first-order autocorrelation of 0.984 were more representative for generating expected future spot rates, the ratio would be only 1.37. Matching the observed volatility therefore requires either choosing the proper long memory properties for the spot rate or adding the evidence on a time-varying risk premium, or both.

It is difficult to discriminate among alternative views of the long memory properties of the spot rate. A standard augmented Dickey-Fuller (ADF) test does not reject the unit root hypothesis for the spot rate and long-term yields. Sample autocorrelations of interest rate levels decrease very slowly. For example, for the spot rate, the 84th order autocorrelation is still 0.395. This slow decay indicates long memory dynamics. Indeed the semiparametric estimates in table 1 indicate that  $d_r \approx 0.9$  and significantly larger than  $1/2$  at a 5% level.

Some predictability of excess returns is evident from the significantly positive first-order autocorrelations. Following the literature, we consider spreads as explanatory variables for the risk price  $\lambda_t$ . Estimates for the integration order of the five-year spread in table 1 are  $d \approx 0.4$ . The point estimate for the ten-year spread is slightly above  $1/2$  but not significantly so. The estimates of the fractional order of spreads motivate a term structure model that allows for stationary, but not necessarily  $I(0)$ , dynamics for the price of risk.

### IV. Models

For both the spot rate as well as the price of risk, we will make specific parametric assumptions to obtain the factor loadings  $b_n$  in equation (10).

#### A. Alternative Spot Rate Models

In a single-factor model the short-rate process, equation (1), has scalar coefficients  $c_j$ , while the innovations  $\epsilon_t$  have a scalar variance  $\sigma^2$ . We will compare three different ARFIMA specifications that differ by the order of integration for the spot rate. The first model is a fractionally integrated process. We add some transitory dynamics for estimation purposes and to give all models the same number of parameters. Consistent with the small, positive autocorrelation in the  $\Delta r_t$  series (see table 1), it appears that a first-order autoregressive term is sufficient for the transitory dynamics. The spot rate model thus becomes

$$(1 - \nu L)(1 - L)^{d_r-1} \Delta r_t = \epsilon_t. \quad (11)$$

The level of the spot rate is fractionally integrated with fractional parameter  $d_r \in (0, 1)$ . The model's impulse responses are given by  $c_j = \sum_{k=0}^j \nu^k (-1)^{j-k} \binom{1-d_r}{j-k}$ . The cumulative impulse responses  $C_{n-1}$  that we need for the term structure factor loadings are of order  $O(n^{d_r})$ . Since we evaluate the model using maturities  $k$  and  $m = 2k$ , a large  $k$  approximation for the volatility ratio  $\mathcal{M}_0$  under the expectations hypothesis gives  $C_{m-1}/C_{k-1} \approx (m/k)^{d_r} = 2^{d_r}$ .

<sup>6</sup> hdl.handle.net/1902.1/FZLJAXFHBW.

<sup>7</sup> www.federalreserve.gov/econresdata/researchdata/feds200628.xls.

<sup>8</sup> research.stlouisfed.org/fred2/series/DTB3.

For the second model, we consider an  $I(1)$  specification that can mimic the properties of the fractional model. In particular, we consider spot rate predictions that are a weighted average of a random walk and a stationary AR(1) with parameter  $\rho$ . These predictions are

$$E_t[r_{t+j}] = \alpha r_t + (1 - \alpha)(\mu_r + \rho^j(r_t - \mu_r)), \quad (12)$$

with  $\alpha$  the weight given to the random walk forecasts and  $\mu_r$  the unconditional mean under the stationary model. This mixture model can be interpreted as a crude approximation of a long memory process. Granger (1980) proves that a mixture of AR(1) processes with random coefficients converges to a fractional process. The process specified here has just two elements instead of the beta-weighted continuum in Granger (1980). In this model,  $c_0 = 1$  and  $c_j = (1 - \alpha)\rho^j(1 - \rho^{-1})$  for  $j > 0$ . With a non-zero weight for the random walk, this is an  $I(1)$  process, which can be written in the ARFIMA(1,1,1) representation:

$$\Delta r_t = \rho \Delta r_{t-1} - (\rho + (1 - \alpha)(1 - \rho))\epsilon_{t-1} + \epsilon_t. \quad (13)$$

If either  $\rho \uparrow 1$  or  $\alpha \uparrow 1$ , the other parameter becomes unidentified. For  $\alpha$  or  $\rho$  close to 1, the MA root almost cancels either the first-difference operator or the AR root. We then have a process that is borderline between  $I(0)$  and  $I(1)$ .

Our third model is the stationary AR(2) process  $(1 - v_1 L - v_2 L^2)(r_t - \mu_r) = \epsilon_t$  for the level of the spot rate with autoregressive parameters  $v_1$  and  $v_2$ . Since for large  $n$ , the impulse responses  $C_{n-1}$  converge to a constant, a stationary AR model will have problems to match the observed increasing volatility of long-term excess bond returns (unless the largest AR root is very close to one).

All three time series models have two parameters for the dynamics, which for future reference we generically denote by  $\theta_p$ .

### B. The Price of Risk

For the time series process of  $\lambda_t$ , we specialize equation (3) to the univariate process,

$$\sigma^2(\lambda_t - \mu_\lambda) = \xi \sum_{j=0}^{\infty} f_j \epsilon_{t-j}, \quad (14)$$

where the  $f_j$  are scalars, normalized by  $f_0 = 1$ , and where  $\xi$  is a scalar parameter that determines both the volatility of the risk premium and the sign of the covariance between  $r_t$  and  $\lambda_t$ . If  $\xi = 0$ , the risk premium is constant.

We specify a fractional model for  $\lambda_t$ , with a lower order of integration than the spot rate. Since all three spot rate models restrict  $d_r \in [0, 1]$ , the term structure model implies that the spread has the same order of integration as the price of risk. For the price of risk, we need a parsimonious model, since the parameters must be estimated from the cross section of

long-maturity bond returns. Since we use only two long-term maturities, we consider the pure fractional  $I(d_\lambda)$  model,

$$(1 - L)^{d_\lambda} \sigma^2(\lambda_t - \mu_\lambda) = \xi \epsilon_t, \quad (15)$$

with  $d_\lambda \in [0, 1/2)$ . The two risk parameters together form the vector  $\theta_R = (d_\lambda, \xi)'$ .<sup>9</sup>

Given our specification, the variance of the price of risk is

$$\text{var}[\sigma^2 \lambda_t] = \sigma^2 \xi^2 \omega^2, \quad (16)$$

with  $\omega^2 = \Gamma(1 - 2d_\lambda)/(\Gamma(1 - d_\lambda))^2$ .

The decomposition in equation (9) splits excess returns in a predictable component  $\sigma^2 \lambda_t$  and an unpredictable shock  $-\epsilon_{t+1}$ . In the ideal case that  $\lambda_t$  is observed, the share of the predictable component in the total variance of excess returns is the predictive  $R^2$ ,

$$R^2 = \frac{\xi^2 \omega^2}{1 + \xi^2 \omega^2}. \quad (17)$$

This is also the maximum  $R^2$  that can be obtained from predictive regressions of excess returns on observed variables such as yield spreads and macroeconomic variables.

## V. Moment Conditions

We estimate the parameters  $\theta_p$ ,  $\theta_R$ , and  $\sigma^2$  by the generalized method of moments (GMM) exploiting three types of moment conditions. The first set are moment conditions for the time series dynamics of the spot rate process and involve only the spot rate parameters  $\theta_p$ . The second type focuses on the volatility of excess returns. These moments impose conditions on the price of risk such that the model can fit the volatility of excess returns of bonds with long maturities. The third type of moment conditions exploits the predictability of excess returns implied by a time-varying price of risk. These conditions add testable overidentifying information.<sup>10</sup>

### A. Spot Rate Moments

Our GMM moment conditions for the time series parameters  $\theta_p$  of the three alternative spot rate models are the least-squares orthogonality conditions

$$E \left[ \epsilon_t \frac{\partial \epsilon_t}{\partial \theta_p} \right] = 0. \quad (18)$$

Equation (18) provides a moment equation for each element in  $\theta_p$  and therefore exactly identifies the time series parameters. For the AR(2) model, the moment conditions involve

<sup>9</sup> In an earlier draft, we also considered the AR(1) specification  $(1 - \varphi L)\sigma^2(\lambda_t - \mu_\lambda) = \xi \epsilon_t$ , which did not give any qualitatively different results.

<sup>10</sup> We do not consider restrictions on constant terms. Unconditional means of excess returns and spreads are unrestricted and estimated by the sample average. To avoid cluttered notation, we omit all constant terms in the moment conditions developed below—for example, we write  $s_t^{(n)}$  instead of  $s_t^{(n)} - \mu_{s^{(n)}}$ .

only the first two autocorrelations of the level of the spot rate. For both the fractional model and the  $I(1)$  mixture model, we implement the moment conditions as proposed in Beran (1995) using the infinite-order autoregressive representation for the first differences of the spot rate. All presample values for  $\Delta r_t$  are set to 0.

The volatility of the spot rate can be estimated from the additional moment condition  $E[\epsilon_t^2] = \sigma^2$ . The univariate prediction errors likely overestimate the spot rate variance if expectations are formed on a broader information set than lagged short-term interest rates alone. We will therefore mostly omit this moment condition and estimate the volatility from the variance of the excess returns on the long-term bonds.

### B. Volatility

The factor structure implies that the unconditional covariance matrix of our two long-term excess returns  $\mathbf{rx}_t = (rx_t^{(k+1)} \quad rx_t^{(m+1)})'$  can be written as

$$\mathbf{V}_0 = E[\mathbf{rx}_t \mathbf{rx}_t'] = \sigma^2 (1 + \xi^2 \omega^2) \mathbf{bb}', \quad (19)$$

where  $\mathbf{b} = (b_k \quad b_m)'$  is the vector of factor loadings. The factor loadings are a function of both the spot rate parameters  $\theta_P$  and the risk parameters  $\theta_R$ . The growth in factor loadings of long-maturity bonds is the quantity most affected by the long memory properties of the spot rate and therefore directly related to the volatility puzzle. The scaling factor  $\sigma^2(1 + \xi^2 \omega^2)$  is the unconditional variance of the factor  $(-\epsilon_t + \sigma^2 \lambda_{t-1})$ . Ideally  $\sigma^2$  should be equal to the variance of the spot rate prediction errors.

The moment conditions can be rewritten in terms of the volatility of a  $k$ -period bond relative to the volatility of an  $m$  period bond,

$$\mathbf{V}_0 = S_k^2 \begin{pmatrix} 1 & \mathcal{M}_0 \\ \mathcal{M}_0 & \mathcal{M}_0^2 \end{pmatrix}, \quad (20)$$

where  $S_k^2 = b_k^2 \sigma^2 (1 + \xi^2 \omega^2)$  is the unconditional variance of the  $k$ -period excess returns and  $\mathcal{M}_0 = b_m/b_k$  is the relative volatility of the returns of the two long-term bonds. The three moment conditions in  $\mathbf{V}_0$  are related to two functions of the parameters:  $S_k^2$  and  $\mathcal{M}_0$ . Since we take the volatility  $\sigma^2$ , and hence  $S_k^2$ , as a free parameter, the relative volatility  $\mathcal{M}_0$  is the only function of the risk parameters identified from the covariance matrix. It therefore imposes a condition on the risk parameters but does not fully identify  $\theta_R$ .

### C. Predictability

A time-varying price of risk implies predictability of excess returns. As a simple measure of predictability, we include the first-order autocovariances of the excess returns. The autocovariances of excess returns can be computed from the factor model (9) as

$$\begin{aligned} \mathbf{V}_1 &= E[\mathbf{rx}_t \mathbf{rx}_{t-1}'] \\ &= \sigma^2 \left( -\xi + \frac{d_k}{1 - d_k} \xi^2 \omega^2 \right) \mathbf{bb}' = \mathcal{M}_1 \mathbf{V}_0. \end{aligned} \quad (21)$$

The four moment conditions in equation (21) identify the first-order autocorrelation,

$$\mathcal{M}_1 = \frac{-\xi + \xi^2 \omega^2 \frac{d_k}{1 - d_k}}{1 + \xi^2 \omega^2}, \quad (22)$$

as one additional function of the risk parameters. Hence, the autocorrelation  $\mathcal{M}_1$  and volatility  $\mathcal{M}_0$  jointly just identify the risk parameters. Due to the factor structure, the autocorrelation does not depend on maturity and/or the parameters of the spot rate process. The maturity independence of  $\mathcal{M}_1$  adds useful overidentifying restrictions to obtain precise estimates of  $\mathcal{M}_1$  and, hence, the risk parameters  $\theta_R$ .

With the moment conditions discussed so far, all spot rate models will have the same fit for the excess return moment conditions. To discriminate among different spot rate models, we consider overidentifying restrictions implied by predictive regressions of the form

$$rx_{t+1}^{(n+1)} = \psi_{0,n} + \psi_n s_t^{(n)} + \eta_{t+1}^{(n+1)}, \quad (23)$$

where excess returns are regressed on their own lagged spread. We refer to the  $\psi_n$  coefficients as the Campbell-Shiller slopes after Campbell and Shiller (1991) who find that regression estimates of  $\psi_n$  are typically positive. Since the term structure model completely specifies the dynamics for excess returns and yield spreads, we can derive the model-implied slopes for the predictive regressions as a function of both  $\theta_P$  and  $\theta_R$  (see appendix C). The Campbell-Shiller regressions thus provide the two additional moment conditions:

$$E \left[ \left( rx_{t+1}^{(n+1)} - \psi_n s_t^{(n)} \right) s_t^{(n)} \right] = 0, \quad n = k, m. \quad (24)$$

Comparing model implied and unrestricted regression slopes has been used as a powerful empirical test of term structure models (see, e.g., Dai & Singleton, 2002, and Sangvinatsos & Wachter, 2005).

A related predictive regression has been proposed by Cochrane and Piazzesi (2005). Instead of regressing each excess return on its own spread  $s_t^{(n)}$ , they construct a common predictive factor  $w_t$  as a linear combination of spreads with different maturities. Having the same explanatory variable for all excess returns is motivated by the factor model (9). The Cochrane-Piazzesi slopes imply the two moment conditions

$$E \left[ \left( rx_{t+1}^{(n+1)} - \beta_n w_t \right) w_t \right] = 0, \quad n = k, m. \quad (25)$$

Expressions for the implied Cochrane-Piazzesi slopes  $\beta_n$  can be derived in the same way as the Campbell-Shiller slopes  $\psi_n$ . Theorem 2 implies that the predictable component has the same factor structure as the shock to excess returns.

In equation (25) we therefore have  $\beta_m = \mathcal{M}_0\beta_k$ , meaning that the two Cochrane-Piazzesi slopes add one new function of the parameters and one overidentifying condition on the relative volatility  $\mathcal{M}_0$ .

For our data, we define  $w_t = s_t^{(k)} - \gamma s_t^{(m)}$  as a linear combination of the five- and ten-year spreads. We fix  $\gamma = 0.54$ , which is the estimate of  $\gamma$  obtained from a nonlinear bivariate OLS regression of five-year and ten-year excess returns on  $w_t$ . The value for  $\gamma$  implies that the prediction factor has tent-shaped weights as in Cochrane and Piazzesi (2005): negative for the spot rate and the ten-year yield and positive for the five-year yield.

#### D. Measurement Error

The single-factor model imposes perfect correlation between the two returns and hence a singular covariance matrix  $\mathbf{V}_0$ . The correlation between the two excess returns is high (0.92) but not perfect. To avoid this stochastic singularity, it is customary to add measurement noise if the number of maturities used for estimation exceeds the number of factors.<sup>11</sup> In line with the literature, we assume that observed yields contain a (small) measurement error  $u_t^{(n)}$ . We assume that the measurement errors follow an AR(1) process with variance  $\zeta_0^2$  and autocorrelation  $\zeta_1$  and are uncorrelated across maturities. The measurement error affects not only  $\mathbf{V}_0$  but also the predictability moment conditions. (Further implementation details are in appendix D.)

## VI. Results

### A. Spot Rate Dynamics

We first look at the implications of different models on long-term expectations for the spot rate. Parameter estimates for all three spot rate models are in table 2. We estimate them by matching all moment restrictions; an alternative set of estimates is obtained from the time series moment conditions, equation (18), only. The two sets of point estimates are very similar; only standard errors differ. The residual variances of the spot rate errors  $\sigma^2$  for the three models are very close to each other. For short-term predictions, they fit the data about equally well.

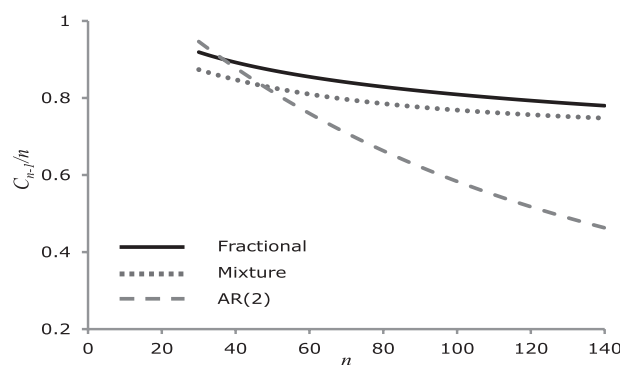
The parameter estimates of the fractional parameter  $d_r$  are around 0.9 and significantly larger than 1/2. They are very close to the semiparametric estimates in table 1. With only spot rate moments, we cannot reject that  $d_r$  is equal to unity. Using all thirteen moments, we find that the overidentifying information from the term structure leads to a much smaller standard error, such that  $d_r$  becomes significantly less than 1. Estimates for the  $I(1)$  mixture model give between 58% and 69% weight to the random walk forecasts and put the remaining weight on a stationary AR(1). The random walk

TABLE 2.—SPOT RATE MODELS

	Time Series			All Moments		
	$d_r$	$\nu$	$\sigma$	$d_r$	$\nu$	$\sigma$
Fractional	0.891 (0.138)	0.237 (0.208)	0.0396 (0.0031)	0.888 (0.045)	0.232 (0.096)	0.0396 (0.0031)
	$\alpha$	$\rho$	$\sigma$	$\alpha$	$\rho$	$\sigma$
Mixture	0.581 (0.533)	0.977 (0.035)	0.0399 (0.0030)	0.693 (0.303)	0.960 (0.064)	0.0399 (0.0030)
	$\nu_1$	$\nu_2$	$\sigma$	$\nu_1$	$\nu_2$	$\sigma$
AR(2)	1.120 (0.096)	-0.134 (0.097)	0.0395 (0.0037)	1.153 (0.095)	-0.167 (0.095)	0.0395 (0.0031)

The table reports GMM estimates of the parameters of three time series models for the three-month T bill rate. The first set of estimates is based on only the spot rate moments, equation (18); the second set uses all thirteen moment conditions. The spot rate volatility  $\sigma$  is estimated from the squared prediction errors of the spot rate and expressed in percent per month. Newey-West standard errors based on six lags are reported in parentheses.

FIGURE 1.—SPOT RATE DYNAMICS



The figure shows the cumulative impulse responses  $C_{n-1}$  scaled by horizon  $n$  for each of the three estimated spot rate models in table 2. Parameters are those obtained using all thirteen moment conditions.

weight  $\alpha$  is, however, estimated with low precision. Long-term forecasts are thus subject to large standard errors. For the AR(2) model, the largest root is 0.983.

What matters for the term structure models are the long-term cumulative impulse responses  $C_{n-1}$  defined in equation (10). Figure 1 shows the sequences  $C_{n-1}$  scaled by  $n$ ; these represent the effect of a shock  $\epsilon_t$  on the expected average future spot rate  $(1/n) E_t [\sum_{j=0}^{n-1} r_{t+j}]$ . For  $n$  between 60 and 120, the sequence of  $C_{n-1}/n$  is largest for the fractional model. Consistent with the estimated  $d_r$ , the decline is very slow. Over the relevant range for  $n$ , the impulse responses for the mixture model are not far from the fractional model. They are a bit below and decline marginally slower. For the mixture model, the sequence  $C_{n-1}/n$  will converge to the random walk weight  $\alpha$ , whereas it will slowly go to 0 at rate  $n^{d_r-1}$  for the fractional model. For the stationary AR(2) model, the scaled cumulative impulse responses decline much faster at rate  $1/n$ .

The implied volatility ratios  $C_{m-1}/C_{k-1}$  are 1.37 for the stationary AR(2) and 1.86 for the fractional and mixture models.<sup>12</sup> For comparison, the volatility ratio of actual long-term bond returns in table 1 is 1.68. Since standard errors on

<sup>11</sup> See, e.g., Campbell and Viceira (2001), de Jong (2000), Christensen et al. (2011), and Duffee (2011).

<sup>12</sup> The transitory dynamics in the fractional model have a negligible impact on the volatility ratio. The large  $k$  approximation  $C_{2k-1}/C_{k-1} \approx 2^{d_r} = 1.86$  is identical to the exact value.



TABLE 3.—RISK PARAMETERS

	$d_\lambda$	$\xi$	$\sigma$	$R^2$	$\zeta_0$	$\zeta_1$
Fractional	0.284 (0.127)	-0.123 (0.041)	0.044 (0.009)	0.019	0.006 (0.001)	0.15 (0.18)
Mixture	0.286 (0.390)	-0.120 (0.045)	0.046 (0.016)	0.018	0.005 (0.002)	0.01 (0.43)
AR(2)	0.498 (0.002)	0.049 (0.029)	0.025 (0.004)	0.146	0.024 (0.004)	0.94 (0.03)

The table reports GMM estimates of the term structure model parameters under alternative assumptions on the spot rate model. Estimates are based on all thirteen moment conditions in section V using a diagonal weighting matrix.  $\sigma$  is estimated from the excess return moments.  $R^2$  is the maximum predictability defined in equation (17) as a function of  $d_\lambda$  and  $\xi$ . Newey-West standard errors with six lags are reported in parentheses.

the long-horizon impulse responses are large for most of the estimates, we cannot reject the hypothesis that  $C_{m-1}/C_{k-1} = 1.68$  for most estimates of the spot rate dynamics. The only exception is the fractional model with parameters estimated from all moment conditions.

The impulse responses illustrate the volatility puzzle within an expectations model. Stationary spot rate dynamics do not produce sufficiently volatile future expectations to explain the volatility of long rates. Conversely, the fractional and mixture models generate too much volatility. Therefore, risk dynamics are needed to increase volatility when the spot rate follows a stationary AR(2), whereas the risk dynamics need to mitigate the expectations effect for the nonstationary models.

### B. Risk Dynamics

Table 3 shows how the risk parameters depend on the alternative spot rate models. As conjectured, the sign of  $\xi$ , the covariance between shocks to the price of risk and the spot rate, depends on the spot rate persistence. For the stationary AR(2) model, the estimate of  $\xi$  is positive; for the two nonstationary spot rate models, the estimate of  $\xi$  is negative.

The sign of  $\xi$  also affects the estimate of  $d_\lambda$ . The expression for  $\mathcal{M}_1$  in equation (22) shows that when  $\xi > 0$ , positive autocorrelation of excess returns is feasible only when the variance of the risk premium becomes extremely large. This explains why the estimated  $d_\lambda$  for the AR(2) model is at the boundary of the stationary region, almost indistinguishable from  $1/2$ . An extremely volatile risk premium is needed to compensate for the low volatility of the expectations part of the term structure. As a result, the AR(2) model implies that the variance of excess returns that can be attributed to the time-varying risk premium, the  $R^2$  in equation (17), is 15%. This is not an artifact of the single-factor model, as Duffee (2011) finds similar high implied predictability within a stationary affine five-factor term structure model. Such a high  $R^2$  is inconsistent with the empirical literature, where the  $R^2$  from predictive regressions with monthly data is mostly around 5%.<sup>13</sup> Duffee (2011) explains why so much volatility in the risk premium is economically implausible as well.

<sup>13</sup> See Fama and Bliss (1987) for an early reference. For our data, we also find a low  $R^2$  for the predictive regressions. The large predictable component in Cochrane and Piazzesi (2005) is for annual returns.

TABLE 4.—MODEL IMPLIED MOMENTS

	$J$	Excess Returns		Campbell-Shiller		Cochrane-Piazzesi	
		$\mathcal{M}_0$	$\mathcal{M}_1$	$\psi_k$	$\psi_m$	$\beta_k$	$\beta_m$
Fractional	3.01 (0.07)	1.752 (0.076)	0.128 (0.043)	2.33 (0.48)	2.59 (0.56)	10.32 (3.12)	18.08 (5.77)
Mixture	2.01 (0.30)	1.766 (0.068)	0.125 (0.043)	2.64 (0.88)	3.26 (0.83)	10.09 (3.01)	17.81 (5.62)
AR(2)	4.97 (0.06)	1.671 (0.059)	0.103 (0.044)	1.89 (0.06)	1.79 (0.05)	35.75 (26.07)	59.72 (45.10)
Unrestricted	0.23 (0.42)	1.666 (0.055)	0.108 (0.043)	3.16 (1.04)	4.29 (1.17)	10.12 (2.93)	16.87 (4.89)

The first column reports the GMM distance  $J$  with the associated Jagannathan-Wang  $p$ -value in parentheses. The next two columns report the implied relative variance ( $\mathcal{M}_0$ ) and autocorrelation ( $\mathcal{M}_1$ ) of five-year ( $k = 60$ ) and ten-year ( $m = 120$ ) excess returns. The right part of the table shows the implied regression coefficients of excess returns on the own maturity spread (Campbell-Shiller) or on the linear combination  $w_t = s_t^{(k)} - 0.54s_t^{(m)}$  (Cochrane-Piazzesi). The final row reports unrestricted moment estimates independent of the term structure parameters (see appendix D). All estimates use the same diagonal weighting matrix. Standard errors in parentheses are derived from the GMM parameter estimates.

Therefore, spot rate dynamics must be (almost) nonstationary to obtain a model with limited volatility for the risk premium. Since  $\xi < 0$  in the fractional and mixture models, matching  $\mathcal{M}_1 > 0$  is easier and a moderate degree of persistence in the price of risk suffices. In the fractional model  $d_\lambda$  is significantly less than  $1/2$ . The term structure estimates for  $d_\lambda$  are thus consistent with the time series evidence in table 1. Estimates for the predictive  $R^2$  in the fractional and the mixture models are much lower than for the AR(2) model and much more in line with evidence from predictive regressions.

Apart from the high implied  $R^2$ , the AR(2) estimates show several other signs of tension. Under a strict interpretation of the single-factor model the estimates for  $\sigma$  in tables 2 and 3 should be equal, whereas they will differ when the spot rate is affected by multiple factors that affect only short- and medium-term maturities. For the AR(2) model, the two estimates are very different, and even significantly so. For the fractional and mixture models, the two estimates for  $\sigma$  are very close, and we cannot reject that the two are equal. Finally, the nonstationary models do not require sizable measurement error. The estimates for  $\zeta_0$  and  $\zeta_1$  imply that measurement error accounts for less than 10% of the fitted variance of excess returns. In contrast, the AR(2) model needs large and persistent measurement error to fit the term structure moment conditions.

### C. Implied Moments

Implied moment estimates are summarized in table 4. The six overidentifying moment conditions are not rejected for any of the three models. This is subject to the earlier observation that the AR(2) model requires a large measurement error component.

For all models, the implied estimate of the volatility ratio  $\mathcal{M}_0$  is close to the unrestricted direct estimate. The volatility ratio is very precisely estimated in the data and therefore has a large weight in the GMM objective function. All models are also able to produce the small, positive autocorrelations for excess returns,  $\mathcal{M}_1$ .

Direct estimates of the Campbell-Shiller and Cochrane-Piazzesi predictive regression slopes are positive, consistent with the literature. Since the standard errors of the regression estimates (reported in the bottom row of table 4) are fairly large, these moments receive less weight in the GMM objective than the volatility moments do. The AR(2) spot-rate model cannot match any of the regression slopes. The implied Campbell-Shiller slopes are around 1.8, which is about half of the values found in the empirical regressions. The model parameters provide sharp estimates of the implied regression slopes, which are, however, far from the unrestricted estimates from the regressions themselves. The Cochrane-Piazzesi slopes are matched even worse: the implied slopes are approximately 3.5 times as high as the ones in the data. Still, the moment conditions are not wildly violated, as the AR(2) model attributes much of the discrepancy between the observed and implied coefficients to a large measurement error.

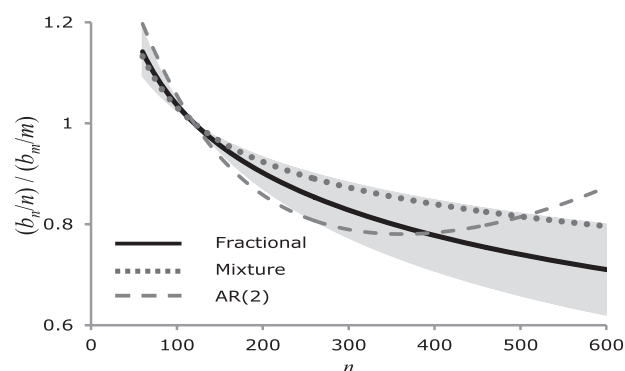
For the fractional and mixture models, the implied regression slopes are closer to the observed ones. Especially the Cochrane-Piazzesi coefficients are almost equal to the actual regression estimates. For the Campbell-Shiller slopes, the empirical regression coefficients remain somewhat larger than what the model is able to generate. Part of this discrepancy may be due to the upward bias of the empirical regression coefficients  $\psi_n$  in small samples.<sup>14</sup>

#### D. Investment Implications

The sign and size of  $\xi$  are important for long-term investors. With a positive  $\xi$  the Merton (1973) model for intertemporal portfolio choice implies that long-term investors should hold positive hedge demands for long-term bonds. The reason is that a positive shock  $\epsilon$  is bad news for current returns, but at the same time, it raises the risk premium, thereby improving future investment opportunities. Within an affine term structure model with a stationary level, such as Sangvinatsos and Wachter (2005), the hedge demand for long-term nominal bonds is therefore positive. The negative value for  $\xi$  associated with the nonstationary spot rate models implies a negative hedge demand. Such negative hedge demands are consistent with empirical models of expected bond returns that do not impose theoretical term structure relations (e.g., Engsted & Pedersen, 2012).

The differences between the alternative models become more pronounced when we extrapolate the factor loadings  $b_n$  to longer maturities. Such an extrapolation is important for the valuation of very long-dated life insurance or pension liabilities. Solvency II regulations stipulate market-consistent valuation. But since markets for very long maturities are less liquid, the long-term discount rates require a model-based extrapolation beyond a certain maturity. Under current Solvency II rules, the yield curve is extrapolated such that the

FIGURE 2.—EXTRAPOLATED YIELD FACTOR LOADINGS



The figure shows the factor loadings  $b_n$  scaled by horizon  $n$ , and relative to the ten-years loading  $b_n/m$  ( $m = 120$ ) for each of the three estimated spot rate models. The shaded area is a  $\pm 1$  standard error band for the fractional model.

market curve is used up to twenty years' maturity, and longer maturity forward rates converge to a constant ultimate forward rate of 4.2% at sixty-years maturity.<sup>15</sup> Converging to a constant in forty years is roughly a mean reversion of 2.5%, which corresponds to an implied volatility of the fifty-year yield that is 21% of the volatility of a ten-year yield.

Our model implies significantly more volatility at the long end. Figure 2 shows the volatility of long-maturity discount rates relative to the ten-year yield. Since for large  $n$ , excess returns are approximately  $n\Delta y_t^{(n)}$ , the volatility of yields behaves as  $b_n/n$ . For the fractional and mixture models, the volatility of the fifty-year rate would be 70% or 80%, respectively, of the volatility of the ten-year rate. Since risk parameters are very similar, the  $I(1)$  mixture model eventually implies larger factor loadings than the  $I(0.9)$  fractional model. Initially the AR(2) curve exhibits the steepest decline, but for maturities beyond thirty years, factor loadings for the AR(2) curve bend upward and diverge.<sup>16</sup> The extrapolation for the AR(2) model is, however, extremely unreliable, with standard errors about four times as large as for the fractional model shown in the figure.

## VII. Conclusion

The properties of the bond risk premium in a term-structure model are sensitive to the persistence in interest rate dynamics. We examine this sensitivity within a generalization of the essentially affine model that allows general linear processes for the factors. While adding modeling flexibility, the model retains an affine structure for excess bond returns and produces a closed-form solution. The model enables a specification where the short-term interest rate is fractionally integrated and nonstationary. The cross-sectional information in the term structure implies that the risk premium

<sup>14</sup> See Bekaert, Hodrick, and Marshall (1997) and Stambaugh (1999) for a detailed analysis of the bias related to the persistence of the spot rate and the strong autocorrelation in the time series of the spread.

<sup>15</sup> The details of the methodology are described in European Insurance and Occupational Pensions Authority (2015).

<sup>16</sup> The upward-sloping part of the AR(2) curve is related to the combination of positive  $\xi$  and the (almost) nonstationary risk premium in this model. The second term in equation (10) then increases at a rate faster than  $n$ .

and the term spread are less persistent. Empirically, such a process can fit many stylized term structure facts. Persistent level shocks can account for the volatility of long-term interest rates, while temporary risk premium shocks capture the predictability of excess returns. The shock has opposite effects on the level and risk premium. An increase in the level reduces the price of risk and expected excess returns.

The nonstationary dynamics for the level of the spot interest rate are contrasted to a stationary autoregressive model. In time series estimation, a stationary AR(2) model cannot be distinguished from a nonstationary model, but the implications for the long end of the yield curve are very different. With stationary spot rate dynamics, expectations of the average future spot rate will have very low volatility. Matching the volatility of long-term interest rates then requires strongly persistent dynamics and implausible volatility for the risk premium. Moreover, the persistence of the spot rate affects the sign of the covariance between a shock to the spot rate level and the risk premium.

For the nonstationary spot rate dynamics we compare two different models. Time series analysis favors a fractionally integrated specification for the short-term interest rate, with a long-memory parameter of  $d \approx 0.9$ . Adding term structure moment conditions leads to a more precise estimate for the fractional parameter. A mixture of a random walk and a stationary AR(1) process appears a close approximation for the fractional process. Both provide similar spot rate expectations over horizons between five and ten years.

## REFERENCES

- Abadir, Karim M., Kaddour Hadri, and Elias Tzavalis, "The Influence of VAR Dimensions on Estimator Biases," *Econometrica* 67 (1999), 163–181.
- Backus, David K., and Stanley E. Zin, "Long-Memory Inflation Uncertainty: Evidence from the Term Structure of Interest Rates," *Journal of Money Credit and Banking* 25 (1993), 681–700.
- Bauer, Michael D., "Term Premia and the News," Federal Reserve Bank of San Francisco working paper 2011-03 (2011).
- Bauer, Michael D., Glenn D. Rudebusch, and Jing Cynthia Wu, "Correcting Estimation Bias in Dynamic Term Structure Models," *Journal of Business and Economics Statistics* 30 (2012), 454–467.
- , "Term Premia and Inflation Uncertainty: Empirical Evidence from an International Dataset: Comment," *American Economic Review* 104 (2014), 323–337.
- Bekaert, Geert, Robert J. Hodrick, and David A. Marshall, "On Biases in Tests of the Expectations Hypothesis of the Term Structure of Interest Rates," *Journal of Financial Economics* 44 (1997), 309–348.
- Beran, Jan, "Maximum Likelihood Estimation of the Differencing Parameter for Invertible Short and Long Memory Autoregressive Integrated Moving Average Models," *Journal of the Royal Statistical Society—Series B* 57 (1995), 659–672.
- Campbell, John Y., and Robert J. Shiller, "Cointegration and Tests of Present Value Relations," *Journal of Political Economy* 95 (1987), 1062–1088.
- , "Yield Spreads and Interest Rate Movements: A Bird's Eye View," *Review of Economic Studies* 58 (1991), 495–514.
- Campbell, John Y., and Luis M. Viceira, "Who Should Buy Long-Term Bonds?" *American Economic Review* 91 (2001), 99–127.
- Chen, Willa W., and Clifford M. Hurvich, "Estimating Fractional Cointegration in the Presence of Polynomial Trends," *Journal of Econometrics* 117 (2003), 95–121.
- Christensen, Jens H. E., Francis X. Diebold, and Glenn D. Rudebusch, "The Affine Arbitrage-Free Class of Nelson–Siegel Term Structure Models," *Journal of Econometrics* 164 (2011), 4–20.
- Cochrane, John H., and Monika Piazzesi, "Bond Risk Premia," *American Economic Review* 95 (2005), 138–160.
- , "Decomposing the Yield Curve," SSRN working paper 1333274 (2008).
- Connolly, Robert A., Z. Nuray Güner, and Kenneth N. Hightower, "Evidence on the Extent and Potential Sources of Long Memory in U.S. Treasury Security Returns and Yields," *Journal of Money Credit and Banking* 39 (2007), 689–702.
- Dai, Qiang, and Kenneth J. Singleton, "Expectation Puzzles, Time-Varying Risk Premia, and Dynamic Models of the Term Structure," *Journal of Financial Economics* 63 (2002), 415–441.
- Dewachter, Hans, and Marco Lyrio, "Macro Factors and the Term Structure of Interest Rates," *Journal of Money Credit and Banking* 38 (2006), 119–140.
- Diebold, Francis X., and Atsushi Inoue, "Long Memory and Regime Switching," *Journal of Econometrics* 105 (2001), 131–159.
- Duffee, Gregory R., "Term Premia and Interest Rate Forecasts in Affine Models," *Journal of Finance* 57 (2002), 405–443.
- , "Information in (and Not in) the Term Structure," *Review of Financial Studies* 24 (2011), 2895–2934.
- Engsted, Tom, and Thomas Q. Pedersen, "Return Predictability and Intertemporal Asset Allocation: Evidence from a Bias-Adjusted VAR Model," *Journal of Empirical Finance* 19 (2012), 241–253.
- European Insurance and Occupational Pensions Authority, "Technical Documentation of the Methodology to Derive EIOPA's Risk-Free Interest Rate Term Structures," EIOPA-BoS-15/035 (2015).
- Fama, Eugene F., "The Information in the Term Structure," *Journal of Financial Economics* 13 (1984), 509–528.
- Fama, Eugene F., and Robert R. Bliss, "The Information in Long Maturity Forward Rates," *American Economic Review* 77 (1987), 680–692.
- Gil-Alana, Luis A., and Antonio Moreno, "Uncovering the US Term Premium: An Alternative Route," *Journal of Banking and Finance* 36 (2012), 1181–1193.
- Granger, Clive W. J., "Long Memory Relationships and the Aggregation of Dynamic Models," *Journal of Econometrics* 14 (1980), 227–238.
- Gürkaynak, Refet S., Brian Sack, and Jonathan H. Wright, "The U.S. Treasury Yield Curve: 1961 to the Present," *Journal of Monetary Economics* 54 (2007), 2291–2304.
- Hamilton, James D., and Jing Cynthia Wu, "Identification and Estimation of Gaussian Affine Term Structure Models," *Journal of Econometrics* 168 (2012), 315–331.
- Iacone, Fabrizio, "A Semiparametric Analysis of the Term Structure of the US Interest Rates," *Oxford Bulletin of Economics and Statistics* 71 (2009), 475–490.
- Jagannathan, Ravi, and Zhenyu Wang, "The Conditional CAPM and the Cross Section of Expected Returns," *Journal of Finance* 51 (1996), 3–53.
- Jardet, Caroline, Alain Montfort, and Fulvio Pegoraro, "No-Arbitrage Near-Cointegrated VAR(p) Term Structure Models, Term Premia and GDP Growth," *Journal of Banking and Finance* 37 (2013), 389–402.
- de Jong, Frank, "Time-Series and Cross Section Information in Affine Term Structure Models," *Journal of Business and Economics Statistics* 18 (2000), 300–314.
- Joslin, Scott, Marcel Pribsch, and Kenneth J. Singleton, "Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks," *Journal of Finance* 69 (2014), 1197–1233.
- Joslin, Scott, Kenneth J. Singleton, and Haoxiang Zhu, "A New Perspective on Gaussian Dynamic Term Structure Models," *Review of Financial Studies* 24 (2011), 926–970.
- Kozicki, Sharon, and Peter A. Tinsley, "Shifting Endpoints in the Term Structure of Interest Rates," *Journal of Monetary Economics* 47 (2001), 613–652.
- Merton, Robert C., "An Intertemporal Asset Pricing Model," *Econometrica* 41 (1973), 867–887.
- Nielsen, Morten Ø., "Nonparametric Cointegration Analysis of Fractional Systems with Unknown Integration Orders," *Journal of Econometrics* 155 (2010), 170–187.
- Sangvinatsos, Antonios, and Jessica A. Wachter, "Does the Failure of the Expectations Hypothesis Matter for Long-Term Investors?" *Journal of Finance* 60 (2005), 179–230.
- Shea, Gary S., "Uncertainty and Implied Variance Bounds in Long-Memory Models of the Interest Rate Term Structure," *Empirical Economics* 16 (1991), 287–312.

- Shiller, Robert J., “The Volatility of Long Term Interest Rates and Expectations Models of the Term Structure,” *Journal of Political Economy* 87 (1979), 1190–1219.
- Shimotsu, Katsumi, “Exact Local Whittle Estimation of Fractional Integration with Unknown Mean and Time Trend,” *Econometric Theory* 26 (2010), 501–540.
- Shimotsu, Katsumi, and Peter C. B. Phillips, “Local Whittle Estimation of Fractional Integration and Some of Its Variants,” *Journal of Econometrics* 130 (2006), 209–233.
- Stambaugh, Robert F., “Predictive Regressions,” *Journal of Financial Economics* 54 (1999), 375–421.
- Sun, Yixiao, and Peter C. B. Phillips, “Understanding the Fisher Equation,” *Journal of Applied Econometrics* 19 (2004), 869–886.
- Wright, Jonathan H., “Term Premia and Inflation Uncertainty: Empirical Evidence from an International Panel Dataset,” *American Economic Review* 101 (2011), 1514–1534.

## APPENDIX A

### Proof of Theorem 1

Start by rewriting the pricing equation (4) in logs, using  $p_t^{(1)} = E_t[m_{t+1}] + \frac{1}{2} \text{var}_t[m_{t+1}]$ , as

$$p_t^{(n+1)} = p_t^{(1)} + E_t[p_{t+1}^{(n)}] + \frac{1}{2} \text{var}_t[p_{t+1}^{(n)}] + \text{cov}_t[m_{t+1}, p_{t+1}^{(n)}]. \quad (\text{A1})$$

To convert from prices to spreads, define the relative price,

$$S_t^{(n)} \equiv -ns_t^{(n)} = p_t^{(n)} - np_t^{(1)}, \quad (\text{A2})$$

and substitute in equation (A1) to arrive at

$$S_t^{(n+1)} = E_t[S_{t+1}^{(n)} + n\Delta p_{t+1}^{(1)}] + \frac{1}{2} \text{var}_t[S_{t+1}^{(n)} + n\Delta p_{t+1}^{(1)}] + \text{cov}_t[m_{t+1}, S_{t+1}^{(n)} + n\Delta p_{t+1}^{(1)}]. \quad (\text{A3})$$

Guess the solution (5) in the text. Initial conditions for  $n = 1$  follow trivially from the definition  $s_t^{(1)} = y_t^{(1)} - r_t \equiv 0$ . For maturities  $n > 1$ , recognize that  $\Delta p_t^{(1)} = -\Delta r_t$ , and evaluate the conditional moments in equation (A3) as

$$E_t[S_{t+1}^{(n)} - n\Delta r_{t+1}] = -a^{(n)} - \sum_{j=0}^{\infty} (d_{j+1}^{(n)} + nc_{j+1})' \epsilon_{t-j}, \quad (\text{A4})$$

$$\text{var}_t[S_{t+1}^{(n)} - n\Delta r_{t+1}] = (d_0^{(n)} + nc_0)' \Sigma (d_0^{(n)} + nc_0), \quad (\text{A5})$$

$$\text{cov}_t[m_{t+1}, S_{t+1}^{(n)} - n\Delta r_{t+1}] = -(d_0^{(n)} + nc_0)' \Sigma \lambda_t. \quad (\text{A6})$$

Substitute these expressions back in equation (A3) and rearrange the terms:

$$\begin{aligned} S_t^{(n+1)} &= -a^{(n)} - \sum_{j=0}^{\infty} (d_{j+1}^{(n)} + nc_{j+1})' \epsilon_{t-j} + \frac{1}{2} (d_0^{(n)} + nc_0)' \Sigma (d_0^{(n)} + nc_0) \\ &\quad - (d_0^{(n)} + nc_0)' \left( \Sigma \mu_\lambda + \sum_{j=0}^{\infty} F_j \epsilon_{t-j} \right), \\ &= - \left( a^{(n)} - \frac{1}{2} (d_0^{(n)} + nc_0)' \Sigma (d_0^{(n)} + nc_0) + (d_0^{(n)} + nc_0)' \Sigma \mu_\lambda \right) \\ &\quad - \sum_{j=0}^{\infty} (d_{j+1}^{(n)} + nc_{j+1} + F_j' (d_0^{(n)} + nc_0))' \epsilon_{t-j}. \end{aligned} \quad (\text{A7})$$

The second line defines the recursion for the coefficients given in equation (6).

## APPENDIX B

### Proof of Theorem 2

To derive the factor structure, first define the innovations

$$S_{t+1}^{(n)} - n\Delta r_{t+1} - E_t[S_{t+1}^{(n)} - n\Delta r_{t+1}] = -(d_0^{(n)} + nc_0)' \epsilon_{t+1}. \quad (\text{B1})$$

Substitute this into the definition for  $rx_{t+1}^{(n+1)}$  from equation (8) using the pricing equation (A1):

$$\begin{aligned} rx_{t+1}^{(n+1)} &= S_{t+1}^{(n)} - S_t^{(n+1)} - n\Delta r_{t+1} \\ &= -(d_0^{(n)} + nc_0)' \epsilon_{t+1} + E_t[S_{t+1}^{(n)} - n\Delta r_{t+1}] - S_t^{(n+1)} \\ &= -(d_0^{(n)} + nc_0)' \epsilon_{t+1} - \frac{1}{2} \text{var}_t[S_{t+1}^{(n)} + n\Delta p_{t+1}^{(1)}] \\ &\quad - \text{cov}_t[m_{t+1}, S_{t+1}^{(n)} + n\Delta p_{t+1}^{(1)}] \\ &= (d_0^{(n)} + nc_0)' (-\epsilon_{t+1} + \Sigma \lambda_t) - \frac{1}{2} (d_0^{(n)} + nc_0)' \Sigma (d_0^{(n)} + nc_0). \end{aligned} \quad (\text{B2})$$

From the last line, we have the definition of the factor loadings as  $b_n = d_0^{(n)} + nc_0$ . This specializes to  $b_1 = c_0$ . For  $n > 1$ , the expression can be simplified using the recursive definition of  $d_0^{(n)}$  in equation (6):

$$\begin{aligned} b_n &= nc_0 + d_0^{(n)} \\ &= nc_0 + (n-1)c_1 + F_0' b_{n-1} + d_1^{(n-1)} \\ &= nc_0 + (n-1)c_1 + (n-2)c_2 + F_0' b_{n-1} + F_1' b_{n-2} + d_2^{(n-2)} \\ &\vdots \\ &= \sum_{i=0}^{n-1} (n-i)c_i + \sum_{i=1}^{n-1} F_{n-i-1}' b_i. \end{aligned} \quad (\text{B3})$$

## APPENDIX C

### Predictive Regressions

We derive the slope coefficients of the regressions of excess bond returns on lagged spreads in equations (24) and (25). The process for the spread is given in theorem 1. For the Campbell-Shiller regression, the slope coefficient follows as

$$\begin{aligned} \psi_n &\equiv \frac{\text{cov}[rx_{t+1}^{(n+1)}, s_t^{(n)}]}{\text{var}[s_t^{(n)}]} \\ &= b_n' \frac{E\left[\left(\sum_{j=0}^{\infty} F_j \epsilon_{t-j}\right) \left(\sum_{j=0}^{\infty} d_j^{(n)'} \epsilon_{t-j}\right)\right]}{E\left[\left(\sum_{j=0}^{\infty} d_j^{(n)'} \epsilon_{t-j}\right)^2\right]} = b_n' \frac{\sum_{j=0}^{\infty} F_j \Sigma d_j^{(n)}}{\sum_{j=0}^{\infty} d_j^{(n)'} \Sigma d_j^{(n)}}. \end{aligned} \quad (\text{C1})$$

For a single-factor model, equation (C1) can be simplified. With  $K = 1$ ,  $\Sigma$  is scalar and cancels in both numerator and denominator, while  $b_n$  and  $F_j = \xi f_j$  are also scalars, leading to

$$\psi_n = b_n \xi \frac{\sum_{j=0}^{\infty} f_j d_j^{(n)}}{\sum_{j=0}^{\infty} (d_j^{(n)})^2}. \quad (\text{C2})$$

The expression for the regression of excess returns on the prediction factor  $w_t = s_t^{(k)} - \gamma s_t^{(m)}$  is very similar. Just replace  $d_j^{(n)}$  by  $\mathbf{d}_j = d_j^{(k)} - \gamma d_j^{(m)}$ , which is independent of  $n$ . The coefficients  $\beta_n$  therefore depend only on  $n$  through the factor loading  $b_n$ .

Analytic expressions for the infinite sums involving fractional impulse responses are cumbersome, with long expressions involving results on the Gaussian hypergeometric function. In practice, we compute the coefficients

using a numerical approximation. When the price of risk is fractional  $I(d)$ , the spread will also be  $I(d)$ . Therefore, the spread coefficients  $d_j^{(n)}$  will converge at the same rate  $j^{d_h-1}$  as the risk coefficients  $f_j$ . Define  $a_j = d_j^{(n)}/f_j$ . We assume that  $a_j = a_{N+1} = a$  is constant for  $j > N$  and compute the slope coefficients as

$$\begin{aligned}\psi_n &\approx b_n \xi \frac{\sum_{j=0}^N a_j f_j^2 + a \sum_{j=N+1}^{\infty} f_j^2}{\sum_{j=0}^N a_j^2 f_j^2 + a^2 \sum_{j=N+1}^{\infty} f_j^2}, \\ &= b_n \xi \frac{a\omega^2 + \sum_{j=0}^N (a_j - a)f_j^2}{a^2\omega^2 + \sum_{j=0}^N (a_j^2 - a^2)f_j^2}.\end{aligned}\quad (\text{C3})$$

The first  $N$  terms of the sum are calculated exactly, while the remainder uses the gamma function result for  $\omega^2$  below equation (16). We set  $N = 1000$ .

## APPENDIX D

### GMM Estimation Details

The complete model consists of thirteen moment conditions for seven parameters: two in  $\theta_p$ , two in  $\theta_R$ , the variance  $\sigma^2$ , and two measurement error parameters.

We assume that observed yields  $\hat{y}_t^{(n)} = y_t^{(n)} + u_t^{(n)}$  contain a measurement error  $u_t^{(n)}$ . The measurement error for the yields implies that measurement error in excess returns is of the form

$$v_{t+1}^{(n+1)} = \hat{r}x_{t+1}^{(n+1)} - rx_{t+1}^{(n+1)} \approx -nu_{t+1}^{(n)} + (n+1)u_t^{(n)}. \quad (\text{D1})$$

We assume that the spot rate  $r_t$  is observed without error. Because of the measurement error, the diagonal elements of  $\mathbf{V}_0$  and  $\mathbf{V}_1$  have the additional

terms  $E[(v_{t+1}^{(n+1)})^2]$  and  $E[v_{t+1}^{(n+1)}v_t^{(n+1)}]$ . Since the measurement error in the yield implies the same measurement error in the spread, the measurement error model also affects the moment conditions (24) and (25) for the predictive regressions.

For the GMM weighting matrix, we define the moment residuals using equation (18) for the spot rate moments evaluated at the ML time series estimates. For the excess return moments, we use the empirical cross products  $rx_t^{(i+1)}rx_{t-\ell}^{(j+1)}$ , while for the predictive regression moments, we use the residuals in equations (24) and (25) with the regression slopes  $\beta_n$  and  $\psi_n$  evaluated at OLS regression estimates. These residuals are demeaned and then used to estimate the moment covariance matrix.

Numerical problems in estimating the parameters of a term structure model are common (see, e.g., Hamilton & Wu, 2012, and Joslin, Singleton, & Zhu, 2011). We also encountered numerical problems in estimating the parameters using an estimate of the optimal weighting matrix. Optimization is numerically stable when we use a diagonal weighting matrix with elements equal to the inverses of the diagonal elements of the residual covariance matrix. With this choice, we obtain convergence with the numerical gradient norm very close to 0 and find the same optimum using different starting values for all runs that converge. To compute standard errors, we estimate the full moment covariance matrix at the final estimates using the Newey-West procedure with automatic bandwidth selection (which results in six lags). Since the diagonal weighting matrix is not optimal, the asymptotic distribution of the GMM distance measure is not the usual chi-squared distribution of Hansen's  $J$ -statistic, but instead is a mixture distribution derived in Jagannathan and Wang (1996, appendix C). We rely on this mixture distribution for testing the overidentifying moment conditions.

For comparing actual and implied moments, we estimate  $\theta_M = (\mathcal{M}_0, \mathcal{M}_1, \psi_k, \psi_m, \beta_k)$  directly as a vector of five free parameters (plus  $\beta_m = \mathcal{M}_0\beta_k$ ) instead of being functions of the term structure parameters  $\theta_p$  and  $\theta_R$ . GMM estimates of  $(\theta_p, \theta_M)$  are based on the same moment conditions and the same weighting matrix as used for  $(\theta_p, \theta_R)$ . They are reported as time series  $(\theta_p)$  in table 2 and unrestricted  $(\theta_M)$  in table 4.